Stochastic Environmental Research and Risk Assessment (2024) 38:1235–1249
https://doi.org/10.1007/s00477-023-02626-7

ORIGINAL PAPER

Multivariate motion patterns and applications to rainfall radar data

Svenja Fischer[1] · Marco Oesting[2] · Alexander Schnurr[3]

Accepted: 23 November 2023 / Published online: 13 December 2023
© The Author(s) 2023

Abstract

The classification of movement in space is one of the key tasks in environmental science. Various geospatial data such as rainfall or other weather data, data on animal movement or landslide data require a quantitative analysis of the probable movement in space to obtain information on potential risks, ecological developments or changes in future. Usually, machine-learning tools are applied for this task, as these approaches are able to classify large amounts of data. Yet, machine-learning approaches also have some drawbacks, e.g. the often required large training sets and the fact that the algorithms are often hard to interpret. We propose a classification approach for spatial data based on ordinal patterns. Ordinal patterns have the advantage that they are easily applicable, even to small data sets, are robust in the presence of certain changes in the time series and deliver interpretative results. They therefore do not only offer an alternative to machine-learning in the case of small data sets but might also be used in pre-processing for a meaningful feature selection. In this work, we introduce the basic concept of multivariate ordinal patterns and the corresponding limit theorem. A simulation study based on bootstrap demonstrates the validity of the results. The approach is then applied to two real-life data sets, namely rainfall radar data and the movement of a leopard. Both applications emphasize the meaningfulness of the approach. Clearly, certain patterns related to the atmosphere and environment occur significantly often, indicating a strong dependence of the movement on the environment.

Keywords Ordinal data analysis · Spatial classification · Limit theorems · Rainfall events

Mathematics Subject Classification 62M10 · 62H20 · 62F12 · 60F05 · 05A05 · 62G30

1 Introduction

The classification of patterns in space is of high importance for many research questions in environmental sciences and engineering. Applications include, but are not limited to: rainfall data, seismic activity, animal movements and air traffic. Spatial classification is usually a demanding task, since it is computationally complex and a huge amount of data has to be collected and processed (Long and Nelson 2013). The problems we have in mind are high dimensional. For example, in our initial example we consider hourly radar rainfall data. At each point in time we have a grid of data, collecting the amount of rainfall at several points in space. A rainfall event or a rainfall cell on these images can be clearly identified visually since it has intensities greater than zero, just as in radar images. Based on these data, we want to classify how the rain cell respectively the 'center' of the cell moves in space. Movement of rainfall cells can, e.g., be caused by wind or orography and it is important to know the movement of the center to obtain information on the general movement and shift of the rainfall cell as well as where the main impact is recorded. Such information can, for example, be used for flood forecasting.

✉ Svenja Fischer
svenja.fischer@rub.de

Marco Oesting
marco.oesting@mathematik.uni-stuttgart.de

Alexander Schnurr
schnurr@mathematik.uni-siegen.de

[1] Department of Civil and Environmental Engineering, Ruhr-University Bochum, Universitätsstr. 150, 44801 Bochum, Germany

[2] Stuttgart Center for Simulation Science (SC SimTech) & Institute for Stochastics and Applications, University of Stuttgart, Allmandring 5b, 70569 Stuttgart, Germany

[3] Department of Mathematics, Universität Siegen, Walter-Flex-Str. 3, 57072 Siegen, Germany

Usually questions of this kind and similar classification demands for spatial data are tackled nowadays using methods from AI (artificial intelligence) or ML (machine learning). For example, machine-learning approaches are used to classify the type of rainfall (Ghada et al. 2022), the covered area of rainfall (Meyer et al. 2016), the variability of rainfall patterns (Ibebuchi and Abu 2023) or animal movement data (Wang 2019).

However, classification via AI, especially deep learning such as deep neural networks, has several drawbacks. First of all, a large amount of training data is needed and sometimes, this is simply not available or requires huge manual efforts. Secondly, if a structural break in the data occurs, it is not possible to adjust a model (since there is no model). Structural breaks in rainfall data may occur, for example, when the recording instruments are changed or updated, when the basis of the calibration of the radar data is changed, the timing of records is altered or more generally if there is a change in climate. Structural breaks in animal movement data may be caused by anthropogenic impacts, such as the construction of roads. The outcome of ML approaches trained with data recorded before the break might therefore not reflect the change or the pre-break data might no longer be relevant for training. Thirdly, some concepts such as deep neural networks are often seen as a kind of black box (Liang et al. 2021), as their many layers are hard to interpret. Recently, many studies focus on how to increase this interpretability, e.g. by adding additional information or so-called explainers (e.g Buhrmester et al. 2021; Liang et al. 2021; Ferreira et al. 2022). A pre-selection of features or a classification of the input data might increase the interpretability.

In the present article we apply so called ordinal patterns to classify high-dimensional motion patterns. Ordinal patterns classify a vector or a time series by the order of each value in the vector. They have the advantage that they are robust against measurement errors and against several kinds of structural breaks (like, e.g. shifts in the data). Furthermore, the algorithms are very quick and the concept is intuitive. Even for data sets with less than 500 data points, one gets reliable results.

Our approach therefore might be a possibility to further contribute to the ongoing process of unboxing the black box of some ML approaches by means of additional information and pre-selection of features (see above). Before applying an ML-algorithms to the data, one at least fixes ‘direction’ or ‘motion’ as *the* feature of interest. ‘The research area of feature extraction or feature learning is concerned with finding efficient mappings from raw data to (low-dimensional) representations that capture information appropriate for the learning task.’ (Mohr 2022, page v). In the present article we describe a method which allows to reduce the data significantly while keeping the relevant information in a context of two-dimensional motion. Let us emphasize that a pre-processing of this kind not only makes the procedure more transparent, but also more efficient (Mumuni and Mumuni 2022).

Different versions of multivariate ordinal patterns have already been used in order to measure the entropy of certain systems (Mohr et al. 2020a) or data sets and in the context of symmetry approximation (Finke et al. 2020).

Here, we use classes of multivariate ordinal patterns in order to analyze the motion of weather phenomena and animals. Analyzing movements of this kind, as well as those of humans or mechanical objects, helps to understand the impact of these phenomena on the study region. For example, movement of rainfall cells can help to identify the most critical rainfall events for a region (namely those that hit the study region for the longest time) and animal movement can provide information on the probability of an animal to occur in a village. But, the classification can also be used as the input for further analyses, such as weather or flood forecast. Currently, easily applicable methods that are able to classify even small data sets (as is frequently the case in environmental research) are often missing. The method presented here closes this gap. We demonstrate that indeed the classification by ordinal patterns delivers meaningful results with respect to the dependence on the spatial resolution and typical rainfall movements in Europe.

Our method can even be generalized to higher dimensions. We provide the mathematical background for such a generalization in Sect. 3. In three dimensions one could think of flying animals or fish in the ocean.

The paper structure is as follows: In the subsequent section we describe the one dimensional case shortly and describe the generalization to multivariate patterns. Section 3 contains the limit theorems for our setting together with a small simulation study using bootstrap. Section 4 finally deals with the data sets for which our methods are designed.

2 From univariate ordinal patterns to multivariate ones

The reader might wonder, why we are starting with the univariate theory. This is due to the following fact: In the first step, our multivariate patterns are vectors of univariate patterns. This means that we are using the classical definitions separately in each component. In the second step, we reduce the massive number of patterns by considering classes of multivariate patterns.

2.1 Univariate patterns

The origin of the whole theory is in the one dimensional framework. This is natural, since the order structure of the data plays an important role. The up-and-down behavior of the time series is nicely captured via ordinal patterns.

Before tackling the multivariate setting, we describe the univariate one: Let $(x_j)_{j=1,\dots,N}$ be a data set with values in $\mathbb{R}$ or $(X_j)_{j\in\mathbb{Z}}$ a univariate time series, that is, a sequence of real-valued random variables X_j defined on the common probability space.

Let us consider a vector $x = (x^1,\dots,x^d) \in \mathbb{R}^d$ which is either a part of our data set within a moving window of length d or a realisation of d consecutive random variables of our time series. We assign the rank of each entry with the entry itself, that is, the vector (1.2, 7, 2) is mapped to (1, 3, 2), since 1.2 is the lowest value (rank 1) and 7 is the highest value (rank 3). The function which assigns the pattern with the vector x is called Π. Mathematically speaking, Π maps d-dimensional vectors into the space S_d of permutations. This space consists of special vectors, which contain each of the values $\{1,\dots,d\}$ exactly once.

By now, we have only considered the case without ties, that is, all values of the vector are different. As soon as the vector x contains a certain value twice (or more often), we call this a tie. In most papers on ordinal patterns it is (sometimes tacitly) assumed, that ties are excluded or at least very rare. Dealing, e.g., with finance data having 4 or more digits, that is a reasonable assumption. However, in climate data one might face data sets with several ties. If ties occur, there are some ways to overcome this problem in the classical setting: either, the definition is slightly modified in a way that still yields a unique pattern. In this case one usually assigns the lower rank with the lower index within the vector, that is, one maps (1, 1, 1) to (1, 2, 3) which is at least counter-intuitive and upward movements are overestimated. Other ways are to exclude the values of the respective window or to randomize in this case. In the latter case one adds a small noise to the sequence which breaks the ties. No matter which procedure is used, information is lost. Hence, as soon as several ties occur, it is useful to follow the approach in Schnurr and Fischer (2022) and consider patterns with ties. In this article, we explicitly allow for ties and assign a larger number of patterns to the vectors under consideration.

Consider again a vector of d consecutive values $(x^1,\dots,x^d)$. Write down the values, which are attained $(y^1,\dots,y^m)$ such that $y^1 < y^2 < \dots < y^m$. Let us emphasize that $m \in \mathbb{N}$ is the number of different values attained in $x^1,\dots, x^d$. Now define the generalized pattern of $(x^1,\dots,x^d)$ to be the vector $\mathbf{t} = (t^1,\dots,t^d) \in \mathbb{N}^d$ such that

$$t^j = k \text{ if and only if } x^j = y^k.$$

We write Ψ for the function which assigns the generalized pattern with each vector in contrast to the Π from above. The set of all generalized patterns of order d is denoted by T_d (in contrast to the classical permutations S_d). The patterns of length $d = 3$ are given in Fig. 1.

A drawback of these generalized patterns is that the number of possible patterns grows even quicker than in the classical case (Table 1).

Let us give an intuitive example for the differences between the classical and the generalized patterns as well as between Π and Ψ: Imagine a race where the time which each runner needed to reach the finish line is sampled in a vector x. If these times are all different to each other, one can use Π in order to assign the rank with each runner. The quicker he/she was, the lower his/her rank. If two or more runners have exactly the same time, Π either randomizes or just assigns the better rank to the person having a smaller starting number. Draws are forbidden. In contrast to this, Ψ allows for draws and results like (1,2,1) are possible.

A comment on the length of the patterns is in place: Experts in the field have emphasized that short pattern lengths like $d = 2$ or $d = 3$ should be used in the context of data analysis (cf. Bandt 2019, and the references given therein). This is due to the fact that each pattern probability is an additional parameter which has to be taken into account and which has to be estimated by statistical methods. In addition, the shorter patterns already contain a lot of information on the longer patterns. As an example consider the case, where one knows the patterns of (x^1, x^2, x^3) and (x^2, x^3, x^4). Then, only very few patterns of length 4 of (x^1, x^2, x^3, x^4) are possible. In some cases it might be only a single possible pattern.

Later, we will use patterns only up to the order 3, which in both cases still yields a tractable number of patterns. The number of classical patterns is just $d!$ while the number of generalized patterns are the so called ordered bell numbers. They can be found, e.g., at the online encyclopedia of integer sequences (oeis.com). There, one also finds several other applications in which the ordered bell numbers appear. To our knowledge, Unakafova and Keller (2013) were the first having calculated these numbers in the context of ordinal patterns.

Analyzing one-dimensional ordinal patterns within a data set, one can describe periodicities, times of monotonicity etc. Furthermore, it is possible to test for serial dependence (Weiß 2022; Weiß and Schnurr 2020), derive the entropy of the underlying system (Piek et al. 2019;

Fig. 1 The 13 generalized patterns of length $d = 3$

up	up at first	constant	down at first	down
(1,2,3)	(1,3,2)		(3,1,2)	(2,1,1)
(1,2,2)	(1,2,1)	(1,1,1)	(2,1,2)	(2,2,1)
(1,1,2)	(2,3,1)		(2,1,3)	(3,2,1)

Table 1 Number of classical and generalized patterns depending on the window length d

d	1	2	3	4	5	6	7
Classical patterns	1	2	6	24	120	720	..
Generalized patterns	1	3	13	75	541	4683	..

Bandt and Pompe 2002) and to calculate the Hurst parameter in the case of long-range dependence (Sinn and Keller 2011).

2.2 Multivariate extensions

Let us now come to the multivariate setting. First we describe what would happen if we tried to generalize the univariate framework into more than one dimension in a naive way. Afterwards we reduce the multivariate complexity in order to derive a tractable and computationally efficient procedure.

As we have pointed out in the introduction, we are considering a 2-dimensional grid with data in $\mathbb{R}$ for every time-point. That is, for every time step $j \in \mathbb{N}$ we have a $k \times \ell$-matrix with values in the real numbers. We consider two standard settings. Either we consider a movement in the points of the grid, i.e., the values that can be attained are in $\{1, ..., k\} \times \{1, ..., \ell\}$, or we consider points in $[1, k] \times [1, \ell]$. An example for the first setting is the movement of the maximum. In this case we track the argmax of the data. The second setting is considered, e.g., if we calculate the center of mass (balance point) which is usually a point not contained in the original grid. While we will usually not obtain ties in the second setting, this is indeed possible in the first one, in particular, if the grid is coarse.

In principle both standard settings can be tackled in the same way, one only has to keep in mind that ties can be a problem (or carry valuable information) in the first setting.

The time series we consider are now $(\boldsymbol{Y}_j)_{j\in\mathbb{Z}}$ having values either in $\{1, ..., k\} \times \{1, ..., \ell\} \subseteq \mathbb{Z} \times \mathbb{Z}$ or in $[1, k] \times [1, \ell] \subseteq \mathbb{R} \times \mathbb{R}$. Our aim is to describe the movement within this derived time series in terms of ordinal patterns.

There have been various attempts in the literature on how to deal with multivariate extensions of ordinal patterns. We give a short overview at the end of this section.

It is possible to consider 3 or more dimensions. However, the number of parameters in the background grows very quickly and it becomes more involved to find classes of patterns sharing interesting properties. Hence, we suggest to use the procedure described here to deal with 2 dimensional data.

Let us consider here explicitly the case of two dimensions. Probably the most natural approach is to consider the vector component-wise and associate with each dimension a pattern as in the one dimensional case (cf. Mohr et al. 2020b, Section 4.1). The two-dimensional pattern is then a vector of patterns or this might be encoded again by a different symbol (which is not canonical, since there is no natural ordering of this two-dimensional structure). A problem, which arises directly—even in dimension 2—is the rapidly growing number of patterns under consideration.

Patterns of length 2 could be considered, but usually they do not carry sufficient information. If we consider patterns of length $d \geq 3$ we have—even in the case without ties—to deal with 36, 576 or 14,400 patterns (for $d = 3, 4, 5$). Higher values are usually not considered in the context of statistics. Let us mention that they are considered in the context of dynamical systems. There, it is even assumed that d tends to infinity (cf. e.g. Keller and Sinn 2010,).

We will (at first) use patterns of order $d = 3$. In all that follows we try to keep the number of objects we consider, and hence the number of parameters, sufficiently low.

Considering patterns of order three in dimension two, we obtain 36 different patterns. In Fig. 2, nine of them are depicted, namely those which start with a movement towards north-east. This is equivalent to the two one-dimensional movements to begin with an upward movement.

The *first idea* is to reduce the number of patterns accordingly to the questions we have in mind and we have pointed out in the introduction. Hence, we use the patterns in order to classify whether we are facing a right turn (R), a left turn (L), monotonicity (M) or inversion (I). The reader might adapt this procedure by making up his/her own classes of patterns depending on a different question or

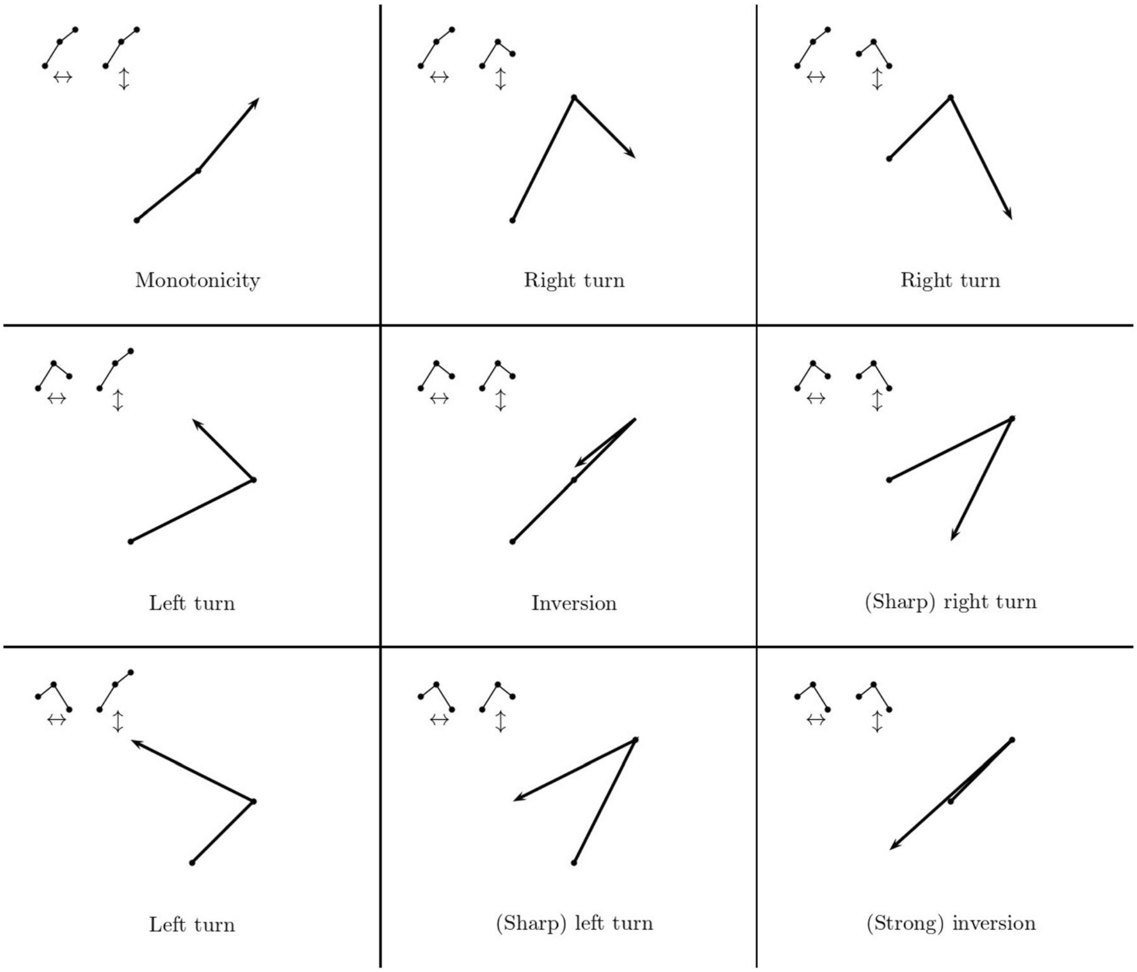

Fig. 2 The nine bivariate patterns moving to north-east first. The two one-dimensional patterns of the components are depicted in the upper left corner of each cell. There, the left one always describes the motion in left-right direction; the right one describes the up-down movement

Table 2 Two-dimensional ordinal patterns of order $d = 3$ assigned to the four classes right turn (R), left turn (L), monotonicity (M) and inversion (I)

R	$\begin{pmatrix}1,2,3\\1,3,2\end{pmatrix}, \begin{pmatrix}3,2,1\\3,1,2\end{pmatrix}, \begin{pmatrix}1,2,3\\2,3,1\end{pmatrix}, \begin{pmatrix}3,2,1\\2,1,3\end{pmatrix}, \begin{pmatrix}1,3,2\\2,3,1\end{pmatrix}, \begin{pmatrix}3,1,2\\2,1,3\end{pmatrix}$
	$\begin{pmatrix}3,1,2\\1,2,3\end{pmatrix}, \begin{pmatrix}1,3,2\\3,2,1\end{pmatrix}, \begin{pmatrix}2,1,3\\1,2,3\end{pmatrix}, \begin{pmatrix}2,3,1\\3,2,1\end{pmatrix}, \begin{pmatrix}2,1,3\\1,3,2\end{pmatrix}, \begin{pmatrix}2,3,1\\3,1,2\end{pmatrix}$
L	$\begin{pmatrix}3,2,1\\1,3,2\end{pmatrix}, \begin{pmatrix}1,2,3\\3,1,2\end{pmatrix}, \begin{pmatrix}3,2,1\\2,3,1\end{pmatrix}, \begin{pmatrix}1,2,3\\2,1,3\end{pmatrix}, \begin{pmatrix}3,1,2\\2,3,1\end{pmatrix}, \begin{pmatrix}1,3,2\\2,1,3\end{pmatrix}$
	$\begin{pmatrix}1,3,2\\1,2,3\end{pmatrix}, \begin{pmatrix}3,1,2\\3,2,1\end{pmatrix}, \begin{pmatrix}2,3,1\\1,2,3\end{pmatrix}, \begin{pmatrix}2,1,3\\3,2,1\end{pmatrix}, \begin{pmatrix}2,3,1\\1,3,2\end{pmatrix}, \begin{pmatrix}2,1,3\\3,1,2\end{pmatrix}$
I	$\begin{pmatrix}1,3,2\\1,3,2\end{pmatrix}, \begin{pmatrix}1,3,2\\3,1,2\end{pmatrix}, \begin{pmatrix}3,1,2\\3,1,2\end{pmatrix}, \begin{pmatrix}3,1,2\\1,3,2\end{pmatrix}, \begin{pmatrix}2,3,1\\2,3,1\end{pmatrix}, \begin{pmatrix}2,3,1\\2,1,3\end{pmatrix}, \begin{pmatrix}2,1,3\\2,3,1\end{pmatrix}, \begin{pmatrix}2,1,3\\2,1,3\end{pmatrix}$
M	$\begin{pmatrix}1,2,3\\1,2,3\end{pmatrix}, \begin{pmatrix}3,2,1\\1,2,3\end{pmatrix}, \begin{pmatrix}3,2,1\\3,2,1\end{pmatrix}, \begin{pmatrix}1,2,3\\3,2,1\end{pmatrix}$

type of analysis. Our limit results remain valid, no matter how classes among patterns are being built.

Let us give an explicit overview on which patterns yield which class. We have L, R, M and I as *pattern classes* in Table 2.

Let us emphasize that instead of dealing with 36 patterns, we deal with 4 pattern classes. In order to derive all patterns which belong to a certain class, the following facts are useful: in one dimension the inverse pattern to (x, y, z) is $(4 - x, 4 - y, 4 - z)$. Inverse means, that one gets the pattern by reflecting it at the x-axis, that is, the up-and-down behavior is completely opposite to the original pattern. Taking one inverse pattern and leaving the pattern in the other dimension as it is, one maps M on M and I on I, left turns become right turns and right turns become left turns. Considering longer time windows means that we get more information on the behavior of the rain area or other phenomena we consider. We could try to use the same procedure for $d = 4, 5, \ldots$ as we have done in the case $d = 3$, that is, we could consider all two-dimensional patterns of that length and then try to classify them. This becomes almost impossible even in the case $d = 4$ (576 patterns). There are patterns having a kind of zik-zak-movement in one dimension, while we have monotonicity in the other dimension and so on.

Hence we make use of our *second idea*: instead of considering longer patterns, we reduce the complexity of the time series again, by only analyzing the behavior in terms of classes L, R, M and I as introduced above. A long left-turn could then be something like L, L, L, that is, 5 bivariate data points which yield a left turn (of any kind) for three consecutive data triplets: (X_j, X_{j+1}, X_{j+2}), $(X_{j+1}, X_{j+2}, X_{j+3})$ and $(X_{j+2}, X_{j+3}, X_{j+4})$ each yield the pattern class L.

The complexity of the time series is again reduced by this procedure, namely we are left with a time series with values in $\{$L, R, I, M$\}$

Let us now analyze what changes, if we consider ties. The main difference is that we can now encounter the absence of a movement in one or two directions, this yields classes of patterns which can be interpreted as a zero movement, the beginning of a motion or the end of a motion. Even in the case $d = 3$ we have $13^2 = 169$ two dimensional ordinal patterns. Combining patterns with ties in one dimension with patterns without ties in the other dimension usually yields a left or right turn. Let us depict nine of the more interesting cases in Fig. 3. The complete overview on the 169 cases is given in Table 7.

In addition to $\{$L, R, I, M$\}$ we have now the total absence of a movement, a zero movement (Z), or a break (B), that is, the motion is stopped. Finally we can observe the start (S) of a motion. The 169 different two dimensional patterns with ties are hence reduced to 7 classes. A data set with values in $\mathbb{R}^{k \times \ell}$ becomes a data set in the finite space $\{$L, R, I, M, B, S, Z$\}$. Admittedly, the three new classes make the procedure (a little) more complicated and we have three more parameters which have to be estimated in applications. However, if ties appear in the data, adding these classes is inevitable. Otherwise we would have to leave out data vectors with ties (which might be a lot) or we would have to randomize as in the one-dimensional case. Either way, we would lose a lot of information.

Having presented our method to deal with multivariate ordinal patterns, let us shortly recall what can be found in the existing literature: Mohr has dealt with different concepts in her doctoral thesis (Mohr 2022). Compare in this context also Mohr et al. (2020b). The different approaches boil down to the following four ideas:

(a) If the data is n-dimensional, consider the ordinal pattern at each fixed point in time. This means that one does *not* use a moving window of length d, but instead compares the values of the different dimensions at each time point separately. The pattern length is n. This approach helps to analyze the dependence between the different components, but does not capture evolution or dependence over time. Compare in this context the spatial approach in Schnurr and Fischer (2022).

(b) Secondly, one could work in a one-dimensional setting in each of the components. That is, one considers the different dimensions separately as it is done, for instance in Oesting and Huser (2022) for longitudinal and latitudinal movements. In the context of permutation entropy one 'pools' afterwards over all dimensions (cf. Keller and Lauffer 2003,). The interplay between the different components is lost completely. Rather, one just glues together the one-dimensional time-series.

(c) Furthermore one could map the multivariate data to scalar data and use classical one-dimension ordinal pattern analysis afterwards. This is done using the concept of depth in Betken and Schnurr (2023). If one uses such a mapping which reduces the dimension, one has to be very careful that no important information on the multivariate structure is lost. The mapping might introduce biases or artifacts that were not present in the original multivariate data set.

(d) Finally, it is possible to consider a vector of patterns for each time window, that is, elements in S_d^2 in the case without ties and elements in T_d^2 in the case with ties. This is done in the first step of our analysis in the present article.

3 Estimation of pattern distributions

In this section, we discuss how to estimate the distribution of multivariate patterns from an excerpt of a stationary time series. Here, we are more general than above and consider arbitrary dimensions p and pattern length d. More precisely, we consider a stationary multivariate time series $\boldsymbol{Y} = (\boldsymbol{Y}_j)_{j \in \mathbb{Z}}$ with values in $\mathbb{Z}^p$ or $\mathbb{R}^p$ ($p \in \mathbb{N}$) and aim at estimating the distribution of multivariate patters $\Pi(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_d)$ and $\Psi(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_d)$, respectively, based on

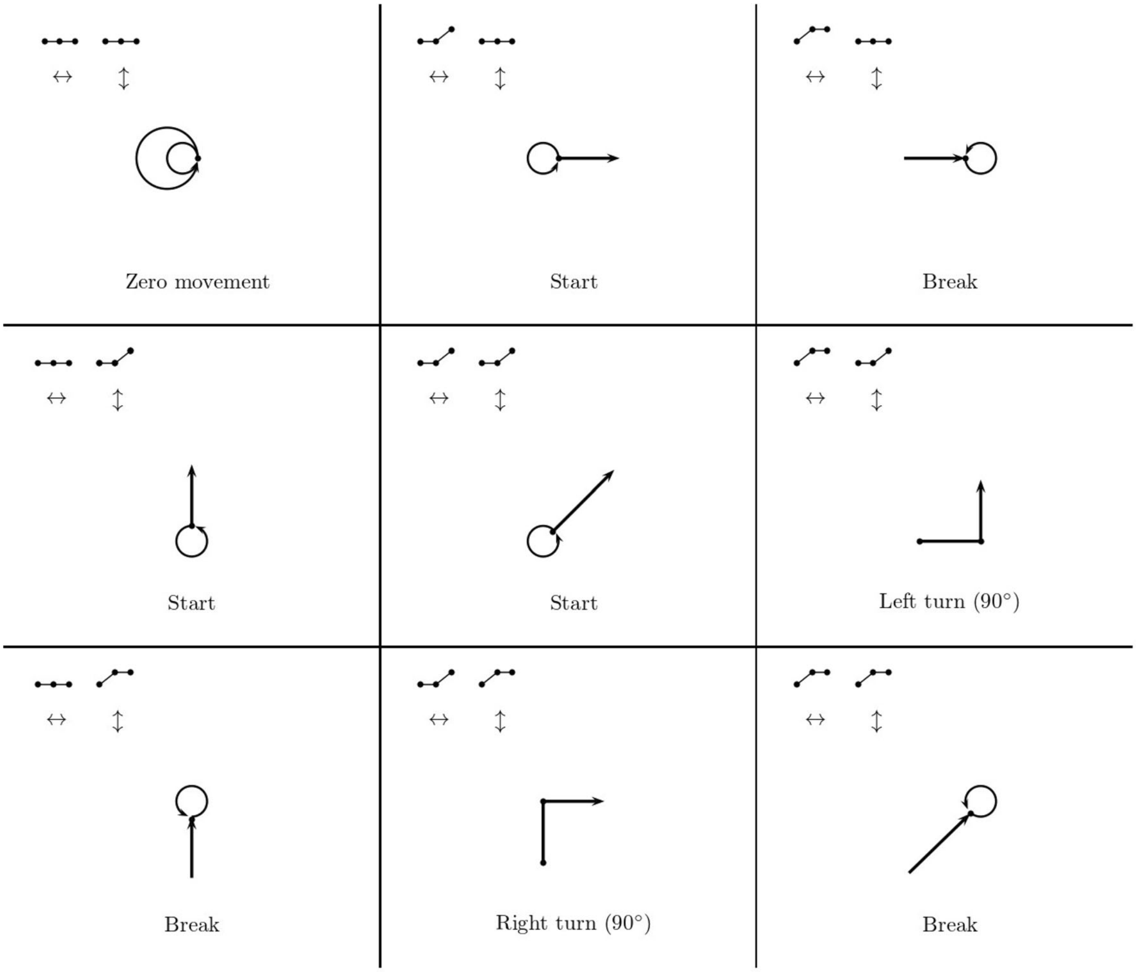

Fig. 3 Nine bivariate patterns with ties moving to north-east. The two one dimensional patterns are again depicted in the upper left corner of each cell

observations of $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_{n+d-1}$. More precisely, we want to estimate $p(A)$ where

$$\begin{cases} p(A) = \Pr(\Pi(\boldsymbol{Y}_1 \ldots, \boldsymbol{Y}_d) \in A), & A \subset S_d^p, \quad \text{if no ties are considered,} \\ p(A) = \Pr(\Psi(\boldsymbol{Y}_1 \ldots, \boldsymbol{Y}_d) \in A), & A \subset T_d^p, \quad \text{if ties are considered,} \end{cases}$$

In particular, this includes the examples for $p = 2$ and $d = 3$ considered in Sect. 2.2.

Due to the stationarity of Y, a natural estimator is given by the sliding window estimator

$$\begin{cases} \widehat{p}_n(A) = \frac{1}{n} \sum_{j=1}^{n} 1\{\Pi(\boldsymbol{Y}_j, \ldots, \boldsymbol{Y}_{j+d-1}) \in A\}, & A \subset S_d^p, \\ \widehat{p}_n(A) = \frac{1}{n} \sum_{j=1}^{n} 1\{\Psi(\boldsymbol{Y}_j, \ldots, \boldsymbol{Y}_{j+d-1}) \in A\}, & A \subset T_d^p, \end{cases}$$

if no ties are considered and if ties are considered, respectively.

In many applications of interest, one might not want to take into account all patterns but only the patterns at times when certain types of events, such as rainfall events, occur. These occurrences are encoded by a binary time series $O = (O_j)_{j \in \mathbb{Z}}$ where

$$O_j = \begin{cases} 1, & \text{if an event occurs at time } j, \\ 0, & \text{if no event occurs at time } j. \end{cases}$$

Assuming stationarity of the $(p+1)$-variate time series $(\boldsymbol{Y}, O) = ((\boldsymbol{Y}_j, O_j))_{j \in \mathbb{Z}}$, we can thus define the estimator

$$\begin{cases} \widehat{p}_n^c(A) = \frac{\sum_{j=1}^{n} 1\{\Pi(Y_j, \ldots, Y_{j+d-1}) \in A, O_j = \ldots = O_{j+d-1} = 1\}}{\sum_{j=1}^{n} 1\{O_j = \ldots = O_{j+d-1} = 1\}}, & A \subset S_d^p, \\ \widehat{p}_n^c(A) = \frac{\sum_{j=1}^{n} 1\{\Psi(Y_j, \ldots, Y_{j+d-1}) \in A, O_j = \ldots = O_{j+d-1} = 1\}}{\sum_{j=1}^{n} 1\{O_j = \ldots = O_{j+d-1} = 1\}}, & A \subset T_d^p, \end{cases} \tag{1}$$

for the conditional pattern distribution

$$\begin{cases} p^c(A) = \Pr(\Pi(Y_1, \ldots, Y_d) \in A \mid O_1 = \ldots = O_d = 1), & A \subset S_d^p, \\ p^c(A) = \Pr(\Psi(Y_1, \ldots, Y_d) \in A \mid O_1 = \ldots = O_d = 1), & A \subset T_d^p, \end{cases}$$

Note that the denominator of the conditional estimator $\widehat{p}_n^c(A)$ might be zero if $O_j = 0$ with positive probability. In this case, it might happen that no sequence of d events in a row is observed and, therefore, the conditional pattern distribution cannot be estimated. If, in contrast, $O_j = 1$ almost surely, the conditional estimator $\widehat{p}_n^c(A)$ coincides with the unconditional estimator $\widehat{p}_n(A)$. Consequently, we henceforth focus on the more general conditional case only.

In the remainder of this section, we will first analyze the asymptotic behaviour of the estimator and establish limit theorems (Sect. 3.1), before demonstrating how to assess the uncertainty of the estimators via bootstrap in numerical experiments (Sect. 3.2).

3.1 Limit theorems

In order to establish asymptotic normality of the estimator, we will have to impose appropriate mixing conditions on the $(p+1)$-variate time series $(\boldsymbol{Y}, O)$. Here, we choose conditions in terms of the strong α-mixing coefficients. For an arbitrary stationary multivariate time series $\boldsymbol{Z} = (\boldsymbol{Z}_j)_{j\in\mathbb{Z}}$, these are defined by

$$\alpha_{\boldsymbol{Z}}(n) = \sup_{A\in\sigma(\boldsymbol{Z}_j : j\leq 0),\, B\in\sigma(\boldsymbol{Z}_j : j\geq n)} |\Pr(A\cap B) - \Pr(A)\Pr(B)|, \quad n = 0, 1, \ldots. \tag{2}$$

The process $\boldsymbol{Z}$ is called strongly α-mixing if $\alpha_{\boldsymbol{Z}}(n) \to 0$ as $n\to\infty$. Examples for processes satisfying this condition include certain Gaussian ARMA processes (see Example 2).

A slightly stronger condition on these mixing coefficients allows us to show asymptotic normality of $\widehat{p}^c(A)$. For notational convenience, the results are stated for the case that ties are considered. However, all the results still hold true in the case of no ties if we replace Ψ by Π and T_d^p by S_d^p, respectively.

Theorem 1 *Let $(\boldsymbol{Y}, O) = ((\boldsymbol{Y}_j, O_j))_{j\in\mathbb{Z}}$ be a stationary $(p+1)$-variate time series with α-mixing coefficients $(\alpha_{(\boldsymbol{Y},O)}(n))_{n=0,1,\ldots}$ satisfying*

$$\sum\nolimits_{k\in\mathbb{N}} \alpha_{(\boldsymbol{Y},O)}(k) < \infty. \tag{3}$$

Then, with $p_d := \Pr(O_1 = \ldots = O_d = 1)$ and for all $A \in T_d^p$, we have

$$\sqrt{n}\begin{pmatrix} \frac{1}{np_d}\sum_{j=1}^n 1\{\Psi(Y_j,\ldots,Y_{j+d-1})\in A, O_j=\ldots=O_{j+d-1}=1\} - p^c(A) \\ \frac{1}{np_d}\sum_{j=1}^n 1\{O_j=\ldots=O_{j+d-1}=1\} - 1 \end{pmatrix} \longrightarrow_d \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2(A) & \sigma_{12}(A)\\ \sigma_{12}(A) & \sigma_2^2(A)\end{pmatrix}\right),$$

where

$$\sigma_1^2(A) = \frac{1}{p_d^2}\sum\nolimits_{j\in\mathbb{Z}} \mathrm{Cov}(1\{\Psi(Y_1,\ldots,Y_d)\in A, O_1=\ldots=O_d=1\}, 1\{\Psi(Y_j,\ldots,Y_{j+d-1})\in A, O_j=\ldots=O_{j+d-1}=1\})$$

$$\sigma_2^2(A) = \frac{1}{p_d^2}\sum\nolimits_{j\in\mathbb{Z}} \mathrm{Cov}(1\{O_1=\ldots=O_d=1\}, 1\{O_j=\ldots=O_{j+d-1}=1\})$$

and $$\sigma_{12}(A) = \frac{1}{p_d^2}\sum\nolimits_{j\in\mathbb{Z}} \mathrm{Cov}(1\{O_1=\ldots=O_d=1\}, 1\{\Psi(Y_j,\ldots,Y_{j+d-1})\in A, O_j=\ldots=O_{j+d-1}=1\}).$$

In particular,

$$\sqrt{n}\left(\widehat{p}_n^c(A) - p^c(A)\right) \longrightarrow_d \mathcal{N}\left(0, \sigma_1^2(A) - 2p^c(A)\sigma_{12}(A) + p^c(A)^2\sigma_2^2(A)\right). \tag{4}$$

The proof of this theorem is given in "Appendix B".

To illustrate the applicability of our result, we consider a flexible class of examples including arbitrary transformations of multivariate Gaussian ARMA models. However, these processes just serve as an example. The class of processes that satisfy the assumptions of Theorem 1 is much richer.

Example 2 For all $j\in\mathbb{Z}$, let the vector $(\boldsymbol{Y}_j, O_j)$ be of the form $f(\boldsymbol{X}_j)$ for some function $f:\mathbb{R}^q\to\mathbb{R}^{p+1}$ and some multivariate stationary Gaussian ARMA process $(\boldsymbol{X}_j)_{j\in\mathbb{Z}}$. Then, we have

$$\boldsymbol{X}_j = H\boldsymbol{Z}_j, \quad j\in\mathbb{Z},$$

where

$$\boldsymbol{Z}_j = F\boldsymbol{Z}_{j-1} + G\boldsymbol{e}_j, \quad j\in\mathbb{Z},$$

for some matrices F, G, H, another stationary time series $(\boldsymbol{Z}_j)_{j\in\mathbb{Z}}$ and i.i.d. random vectors $\boldsymbol{e}_j$, $j\in\mathbb{Z}$, with $\boldsymbol{e}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_r)$. Then, obviously, $\mathbb{E}(\|\boldsymbol{e}_j\|^\delta) < \infty$ and

$$\begin{aligned}
&\int \frac{1}{\sqrt{2\pi}^r}\left|\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\theta})^\top(\boldsymbol{x}-\boldsymbol{\theta})\right) - \exp\left(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{x}\right)\right| \mathrm{d}\boldsymbol{x} \\
&= \mathbb{E}\left|1 - \exp(\langle\boldsymbol{e}_j,\boldsymbol{\theta}\rangle)\exp\left(-\frac{1}{2}\|\boldsymbol{\theta}\|^2\right)\right| \\
&\leq (1+\exp(\|\boldsymbol{\theta}\|^2))\sum_{i=1}^r |\theta_i|\mathbb{E}\left(|e^{(i)}|\exp(|\theta_i e^{(i)}|)\right) \in \mathcal{O}(\|\boldsymbol{\theta}\|_1)
\end{aligned}$$

as $\|\boldsymbol{\theta}\|_1 \to 0$. If the modulus of all the eigenvalues of F is less than 1, by Thm. 3.1 in Pham and Tran (1985), the α-mixing coefficients of $(\boldsymbol{X}_j)_{j\in\mathbb{Z}}$ decay at an exponential rate and so do $|\alpha_{(\boldsymbol{Y},O)}(n)| \leq |\alpha_{\boldsymbol{X}}(n)|$. Thus, the summability condition in Theorem 1 holds.

It has to be noted that even though the underlying process $(\boldsymbol{X}_j)_{j\in\mathbb{Z}}$ is assumed to be Gaussian, the function f allows for very flexible modelling of the process $(\boldsymbol{Y}_j)_{j\in\mathbb{Z}}$. This includes phenomena such as multimodal distributions and heavy tails and thus go beyond the assumption of Gaussianity which is mathematically convenient, but might be unrealistic in practice.

3.2 Assessing the uncertainty of the estimators via bootstrap

While Eq. (1) provides a straightforward point estimate for the probability masses of the conditional pattern distribution, Theorem 1 can be used to assess its uncertainty. The specification of concrete confidence intervals, however, requires knowledge of the asymptotic variance in Eq. (4). As this expression is quite involved and can hardly be estimated directly, it is advisable to resort to resampling techniques such as bootstrap in order to assess the uncertainty of the estimate. Here, it is important to note that our procedure highly relies on the temporal order and temporal dependence structure of the data, which therefore necessarily has to be taken into account by the bootstrap. Thus, we propose the use a multiplier (disjoint) block bootstrap procedure (Bücher and Ruppert 2013; Drees 2015; Bücher and Kojadinovic 2016).

More precisely, we divide our time series into disjoint blocks of length ℓ,

$$\{(O_{(i-1)\ell+1}, \boldsymbol{Y}_{(i-1)\ell+1}), \ldots, (O_{i\ell}, \boldsymbol{Y}_{i\ell})\}, \quad i = 1, \ldots, \lfloor n/\ell \rfloor,$$

estimate both the numerator and denominator in Eq. (1) separately and reweight them by i.i.d. multiplier random variables $1+\xi_1, \ldots, 1+\xi_{\lfloor n/\ell\rfloor}$ satisfying $\mathbb{E}(\xi_i) = 0$ and $\mathrm{Var}(\xi_i) = 1$. Thus, we obtain the bootstrap estimator

$$\widehat{p}_n^{c,B}(A) = \frac{\sum_{i=1}^{\lfloor n/\ell\rfloor}(1+\xi_i)\sum_{j=(i-1)\ell+1}^{i\ell-d+1} 1\{\Pi(Y_j,\ldots,Y_{j+d-1}) \in A, O_j = \ldots = O_{j+d-1} = 1\}}{\sum_{i=1}^{\lfloor n/\ell\rfloor}(1+\xi_i)\sum_{j=(i-1)\ell+1}^{i\ell-d+1} 1\{O_j = \ldots = O_{j+d-1} = 1\}}.$$

By independently sampling new multiplier random variables N_{BS} times, we obtain a sample of independent bootstrap estimates (conditionally on the data) whose distribution can be perceived as an approximation of the distribution of the original estimator.

We illustrate this procedure by a small simulation study. Here, we consider a special case of the multivariate stationary Gaussian ARMA time series in Example 2. More precisely, we consider a trivariate time series $(\boldsymbol{Z}_j)_{j\in\mathbb{Z}}$ of the form:

$$\boldsymbol{Z}_j := \begin{pmatrix} Z_j^{(1)} \\ Z_j^{(2)} \\ Z_j^{(3)} \end{pmatrix} = \begin{pmatrix} \rho_1 Z_{j-1}^{(1)} + \sqrt{1-\rho_1} e_j^{(1)} \\ \rho_2 Z_{j-1}^{(2)} + \sqrt{1-\rho_2} e_j^{(2)} \\ \rho_3 Z_{j-1}^{(3)} + \sqrt{1-\rho_3} e_j^{(3)} \end{pmatrix},$$

where $\boldsymbol{e}_j, j \in \mathbb{Z}$, are i.i.d. with $\boldsymbol{e}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_3)$ and parameters $\rho_1, \rho_2, \rho_3 \in (-1, 1)$. Then, by construction, the stationary solution possesses independent components and standard Gaussian marginal distribution. We set

$$O_j = 1\{Z_j^{(1)} > \Phi^{-1}(0.1)\}, \qquad \boldsymbol{Y}_j = (Z_j^{(2)}, Z_j^{(3)})^\top.$$

Consequently, we expect events occurring in 90 % of all cases.

We consider three different parameter settings:

(i) independence: $\rho_1 = \rho_2 = \rho_3 = 0$
(ii) moderate dependence: $\rho_1 = \rho_2 = \rho_3 = 0.5$
(iii) strong dependence: $\rho_1 = \rho_2 = \rho_3 = 0.8$

In each case, we simulate $N = 100$ time series of length $n = 10\,000$ and estimate the probabilities

$$\widehat{p}_{n,j}^c(L),\ \widehat{p}_{n,j}^c(R),\ \widehat{p}_{n,j}^c(I),\ \widehat{p}_{n,j}^c(M), \quad j = 1, \ldots, N,$$

for patterns of length $d = 3$, where the additional index j refers to the sample under consideration. For each sample, we estimate the uncertainty by the bootstrap sample standard deviation $\hat{\sigma}_j^{(BS)}(A), A \in \{L, R, I, M\}$, based on a block length of $\ell = 100$ and a bootstrap sample size of $N_{BS} = 500$. We then compare the sample standard deviations

$$\widehat{\sigma}(A) = \frac{1}{N-1}\sum_{j=1}^{n}\left[\widehat{p}_{n,j}^c(A) - \overline{p_n^c}(A)\right]^2,$$

where

$$\overline{p_n^c}(A) = \frac{1}{N}\sum_{j=1}^{n}\widehat{p}_{n,j}^c(A), \quad A \in \{L, R, I, M\},$$

to the average bootstrap standard deviations $\overline{\sigma^{(BS)}}(A) = \frac{1}{N}\sum_{j=1}^{n}\hat{\sigma}_j^{(BS)}(A),\ A \in \{L, R, I, M\}$. The results are reported in Table 3. It can be seen that, due to the symmetry of the two components of Y, the pattern class L and R are equally likely in all three scenarios. The monotone patterns M, however, occur less frequently than the inverse patterns I in the independent scenario, while the order of these two frequencies changes in the strongly dependent scenario. In other words: Bivariate movements become more persistent in case of stronger dependence. Due to the large sample size n, the overall uncertainty is small with standard deviations between 0.3 % and 0.7 % which are very well recovered by the bootstrap estimates. We note that these

Table 3 Relative frequencies $\overline{p_n^c}(A)$ of patterns, the empirical standard deviations $\widehat{\sigma}(A)$ and the estimated standard deviations $\overline{\sigma^{(BS)}}(A)$ obtained via bootstrap, $A \in \{L, R, I, M\}$, for the independent, the moderately dependent and the strongly dependent scenario (from top to bottom)

Pattern class	Frequency $\overline{p_n^c}(A)(\%)$	Empirical SD$\widehat{\sigma}(A)(\%)$	Bootstrap SD $\overline{\sigma^{(BS)}}(A)(\%)$
Scenario (i)—Independence			
L	33.3	0.65	0.66
R	33.3	0.69	0.66
I	22.3	0.47	0.50
M	11.2	0.34	0.36
Scenario (ii)—Moderate dependence			
L	32.7	0.59	0.60
R	32.7	0.61	0.59
I	16.9	0.43	0.43
M	17.6	0.43	0.42
Scenario (iii)—Strong dependence			
L	32.0	0.51	0.54
R	32.0	0.48	0.54
I	14.1	0.36	0.39
M	21.9	0.45	0.46

results are invariant under monotone marginal transformations, that is, they also hold true for non-Gaussian processes with the same dependence structure.

4 Application: real life data examples

4.1 Example 1: Rainfall radar data

As an example, we consider rainfall radar data for a catchment in Germany. Rainfall radar in Germany are recorded for more than twenty years and are available in a bias-corrected version from the German Weather Service (DWD), denoted by RADOLAN (Winterrath et al. 2018). Due to their high spatial (1×1 km^2) and temporal (hourly data) resolution, rainfall radar data are used more and more often as input data for hydrological simulation and modelling of runoff. Moreover, they are used as training data sets for short-term flood forecasting based on artificial intelligence. It is therefore of high practical relevance to characterise patterns in rainfall events and their movement in space, e.g. of the focus point. Here, rainfall events are defined as periods of rainfall where the rainfall in one grid cell exceeds the rainfall sum of a given level. We chose the level as the return period of one year of the annual maximum 1 h-rainfall sums, which is an often used input for simulations for the sewer network (Huard et al. 2010). The corresponding rainfall event then is defined as the period where the threshold of 0.1 mm per hour (a typical threshold to differ between wet and dry spells, see, e.g., Huang et al. 2015,) is exceeded in any grid cell. As an example, a catchment in Saxony, eastern Germany is considered. The Tannenberg catchment has a size of 92 km^2, which leads to a consideration of 210 grid cells of radar data (including a 5 km buffer). A total number of 254 events was identified for this catchment in the period 2001–2020. An example of one identified rainfall event is given in Fig. 4.

To track the spot with highest rainfall amount of the rainfall event, the focus point is considered. This focus point is estimated by using the weighted mean of all cells, where the weights are chosen as the rainfall heights of each cell. Hence, as long as the overall rainfall pattern changes (which is usually the case when considering 1-hour time steps), ties will not occur. In such a case, only the four main equivalence classes have to be considered: L, R, M, I. Here, bivariate patterns of length three are considered.

The results for the radar rainfall events in Tannenberg catchment with 1×1 km grid resolution are given in Table 4, together with the theoretical frequency of each equivalence class under independence of both components (directions) and independence in time for each component, which serves as comparison value to evaluate the significance of the results. Additionally, also the frequency of each equivalence class under consideration of short-range dependence (SRD) of each component but independence between the components is considered (see the ARMA model in Sect. 3.2). The results demonstrate that the frequency of monotone movement is much higher than expected, while the left turn movement is lower than expected when compared to completely independent components. This implies that indeed SRD of each component can be expected, as this increases the frequency of monotone behaviour (see last two columns of Table 4). This dependence can be expected when considering hourly rainfall data. Interestingly, also the inverse movement occurs frequently, which is counter-intuitive to what one would expect from rainfall cells. This might be due to the methodology used here to define the focus point. The

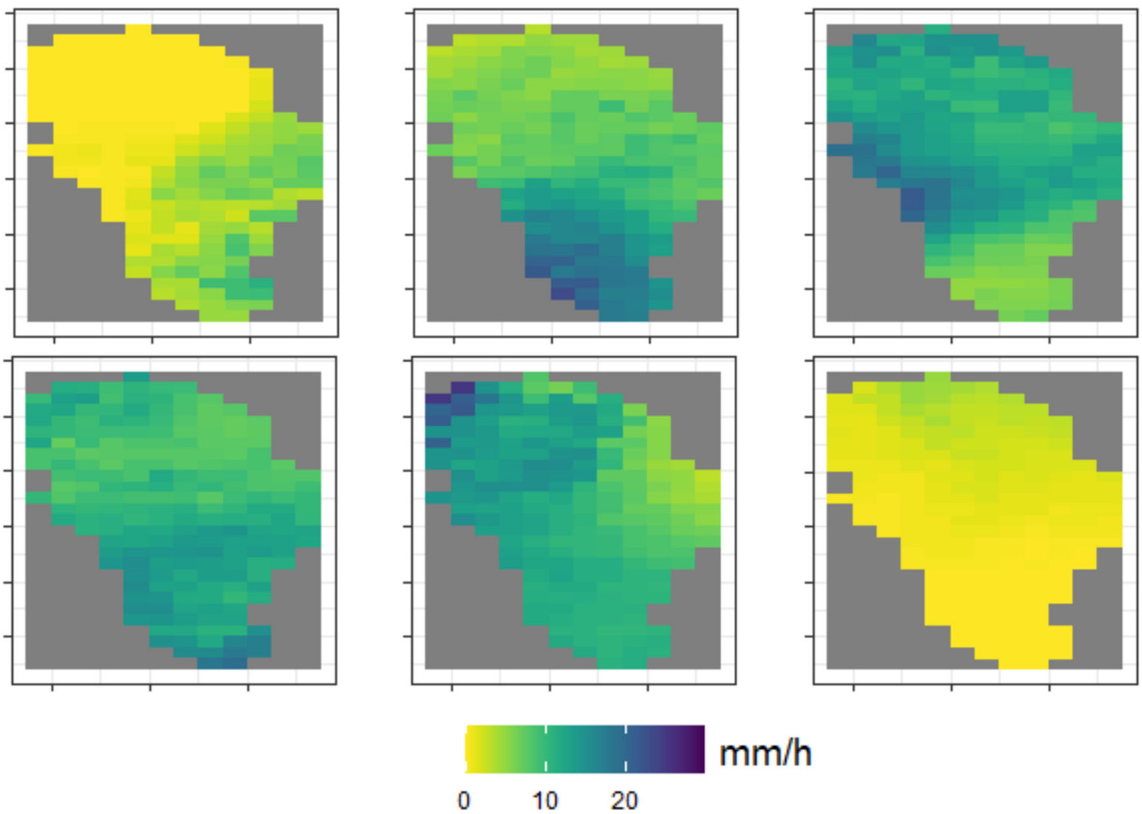

Fig. 4 Example of one six-hour rainfall event for the Tannenberg catchment in Germany

Table 4 Frequency of patterns for the focus point of rainfall events in the Tannenberg catchment in Saxony and theoretical frequency of each equivalence class under independence of all and between all components and under short-range dependence (SRD) in time of each component (in %)

		Theoretical		
Pattern class	Observed Frequency	Independence	SRD ($\rho_i = 0.5$)	SRD ($\rho_i = 0.8$)
L	21.5	33.3	32.7	32.0
R	32.3	33.3	32.7	32.0
I	23.4	22.3	16.9	14.1
M	22.8	11.2	17.6	21.9

applied methodology is sensitive and might lead to changing patterns simply due to a shift in the outer cells of the rainfall event. Yet, this is not what we expect to observe for rainfall events. Therefore, it might be more intuitive to use the argmax-function instead of the focus point to determine the grid cell with maximum rainfall per time step. Yet, this approach also has a drawback: Depending on the spatial resolution of the data, it might occur that the cell with maximum rainfall does not change between two consecutive time steps. This introduces the presence of ties in patterns, and hence the number of equivalence classes that are considered increases, as also the classes Z, S and B have to be taken into account.

To additionally investigate the impact of the resolution of the data, the spatial resolution is varied from small scale (1×1 km) to large scale (4×4 km) such that changes on small distance of 1 km-steps as well as on large distance of 4 km-steps can be considered. The results are given in Table 5 together with the theoretical frequency of each equivalence class under independence, which again serves as comparison value to evaluate the significance of the results. Specific results under SRD are not given here, since the results are crucially influenced by the model choice for the ties, e.g., INAR models or Poisson-increments. However, the same tendency as for previous example can be assumed under dependence of the components.

The results demonstrate that the frequency of the patterns depends much on the spatial resolution of the underlying radar data. While a counter-clockwise movement (L) of the storm cell occurs almost constantly with a frequency nearly equal to the theoretical one, the frequency of storms with clockwise movement (R) decreases rapidly with decreasing spatial resolution and is far below what would have been expected under independence and also below what can be expected under SRD (as also for SRD a similar frequency of L and R patterns should be expected). This implies that for large scale patterns, which are better

Table 5 Frequency of equivalence classes for the maximum of rainfall events in the Tannenberg catchment in Saxony dependent on the spatial resolution of the underlying radar image and theoretical frequencies under independence of all and between all components (in %)

Pattern class	1 × 1 km	2 × 2 km	3 × 3 km	4 × 4 km	theor. independence
L	33.9	33.9	32.3	31.4	35.5
R	35.6	35.2	28.5	20.8	35.5
I	14.2	12.2	12.6	11.7	14.2
M	10.2	5.1	4.9	4.9	4.7
S	2.9	6.2	9.9	13.3	4.7
B	3.1	5.6	10.3	15.0	4.7
Z	0.0	0.0	0.0	0.0	0.6

captured by low-resolution radar patterns as small changes do not disturb the overall storm pattern, counter-clockwise storm movements occur much more frequently than any other pattern. This is line with the fact that in the Northern Hemisphere winds blow counter-clockwise around a center of low pressure, hence coming along with wet conditions and frontal weather system (Coriolis effect). Such a dependence between the components, i.e. the directions, is not considered in the simulation study, which might be the reason why the resulting frequencies differ. Monotone movement on the small scale (i.e., visible in high resolution data) is more frequent than expected, indicating a rapid movement of the storm cells and again short-range dependence of the components. This is meaningful when considering that the focus in this study is laid on heavy rainfall events, i.e., events with high rainfall intensity, where usually fast movements of the rainfall cells can be observed. Significant stops or breaks in the movement only become present when focusing on the large scale, as again here the movement of small rainfall cells is levelled out by the low resolution. Here, the role of ties becomes apparent for this framework: The focus point of a rainfall event does not necessarily have to move between two time steps, as we are using grid cells. Therefore, it is possible that movements only occur within one grid cell and hence are not visible in the data. This might be meaningful when small movements are considered as random and should not impact the resulting pattern. The consideration of ties becomes more important when the spatial resolution decreases, as ties are required when aiming to differ between small and large scale patterns. The consideration of ties thus enables one to consider small as well as large scale changes within the same framework and to omit too sensitive estimation of the focus point.

To test the significance of the results, a second catchment in Bavaria, Germany was also considered. The catchment Baiersdorf with a catchment size of about 70 km^2 is located in the northern parts of Bavaria close to the uplands. Each radar image related to this catchment includes 89 grid cells. Detailed results are not presented here, as these are similar to the ones presented for Tannenberg. Therefore, the overall tendency of the results can be confirmed.

The results obtained here could also be used to identify the most critical patterns of rainfall for a catchment. For this purpose, the patterns of those events could be detected that have the largest overlap with the catchment in time. This is of course highly dependent on the shape of the catchment but can be a valuable information for hydrological modelling of extremes.

4.2 Example 2: Animal movement data

For the second example, we consider data of animal movement. Such data can be expected to have a high spatial resolution as GPS tracker deliver (sub)daily positions of latitude and longitude of the animal's position. Therefore, ties will not occur in the data and no scale-dependence or scale-sensitivity has to be considered. We consider data provided by the online repository Movebank (www.movebank.org), which is coordinated by the Max Planck Institute of Animal Behavior, the North Carolina Museum of Natural Sciences, and the University of Konstanz. There, large data sets on animal tracking are freely available. We consider the data set on daily satellite location data for leopard management in the North West Province, South Africa, 2014–2020 (Power et al. 2021). For the leopard "Brandy" (panthera pardus), we have 451 days of tracking positions for the given time period in a region close to Pretoria, South Africa. The habitat the leopard lives in is framed by cliffs in the south and a highway in the north (Fig. 5).

The high frequency of monotone movement again emphasizes that a dependence in time of each component can be assumed, i.e. the direction of movement of the leopard is continuous for several days (Table 6). Yet, the high frequency of movement towards the left compared to that of movement to the right also indicates that the leopard is impacted by the topography, more precisely the cliff in the south and the highway in the north which forces him to take a left turn, when considering the starting point of the data. A simulation of this movement, e.g., should therefore take into account not only dependence in time but also dependence between the directional components. This can be detected by the proposed approach and used as basis for further analyses.

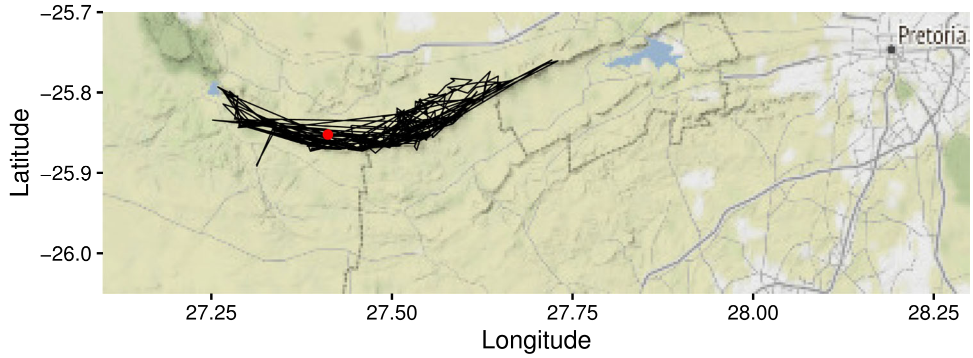

Fig. 5 Daily location data for the leopard "Brandy" in South Africa. The red dot marks the position at the first day of observation

Table 6 Frequency of patterns for the daily location of a leopard in South Africa and theoretical frequencies under independence of all and between all components and under short-range dependence (SRD) in time of each component (in %)

Pattern class	Observed Frequency	Theoretical		
		Independence	SRD ($\rho_i = 0.5$)	SRD ($\rho_i = 0.8$)
L	29.2	33.3	32.7	32.0
R	23.8	33.3	32.7	32.0
I	20.5	22.3	16.9	14.1
M	26.5	11.2	17.6	21.9

Appendix A: Overview of two-dimensional patterns and their equivalent classes

See Table 7.

Appendix B: Proof of Theorem 1

To show the bivariate convergence, by the Cramér-Wold device, it suffices to show that, for every $a, b \in \mathbb{R}$,

$$\frac{1}{\sqrt{n}} \sum\nolimits_{j=1}^{n} Z_j^{a,b} \longrightarrow_d \mathcal{N}\Big(0, a^2\sigma_1^2(A) + 2ab\sigma_{12}(A) + \sigma_2^2(A)\Big), \tag{B1}$$

where

$$Z_j^{a,b} = a\Big[\frac{1}{p_d} 1\{\Psi(Y_j, \ldots, Y_{j+d-1}) \in A, O_j = \ldots = O_{j+d-1} = 1\} - p^c(A)\Big] + b\Big[\frac{1}{p_d} 1\{O_j = \ldots = O_{j+d-1} = 1\} - 1\Big].$$

To this end, we first note that the time series $Z^{a,b} = (Z_j^{a,b})_{j \in \mathbb{Z}}$ is stationary and centered, i.e., $\mathbb{E}(Z_j^{a,b}) = 0$. As each $Z_j^{a,b}$ is a measurable function of $Y_j, \ldots, Y_{j+d-1}$ and $O_j, \ldots, O_{j+d-1}$, we obtain the bound

Table 7 All two-dimensional patterns of length 3 and their respective classes

	(1,2,3)	(1,3,2)	(2,3,1)	(2,1,3)	(3,1,2)	(3,2,1)	(1,2,2)	(1,1,2)	(1,2,1)	(1,1,1)	(2,1,2)	(2,2,1)	(2,1,1)
(1,2,3)	M	L	L	R	R	M	L	R	L	M	R	L	R
(1,3,2)	R	I	L	R	I	L	R	R	R	I	R	L	L
(2,3,1)	R	R	I	I	L	L	R	R	R	I	L	L	L
(2,1,3)	L	L	I	I	R	R	L	L	L	I	R	R	R
(3,1,2)	L	I	R	L	I	R	L	L	R	I	L	R	L
(3,2,1)	M	R	R	L	L	M	R	L	R	M	L	R	L
(1,2,2)	R	L	L	R	R	L	B	R	L	B	R	L	B
(1,1,2)	L	L	L	R	R	R	L	S	L	S	R	S	R
(1,2,1)	R	R	L	R	L	L	R	R	I	I	I	L	L
(1,1,1)	M	I	I	I	I	M	B	S	I	Z	I	S	B
(2,1,2)	L	L	R	L	R	R	L	L	I	I	I	R	R
(2,2,1)	R	R	R	L	L	L	R	S	R	S	L	S	L
(2,1,1)	L	R	R	L	L	R	B	L	R	B	L	R	B

$$\alpha_{Z^{a,b}}(k) \leq \alpha_{(Y,O)}((k-d+1)_+), \qquad n = 0, 1, \ldots.$$

Consequently, Eq. (3) implies that

$$\sum_{k\in\mathbb{N}} \alpha_{Z^{a,b}}(k) < \infty. \tag{B2}$$

In particular, $\alpha_{Z^{a,b}}(n) \to 0$ as $n \to \infty$, which implies ergodicity of $Z^{a,b}$ (see Remark 4.1 in Rio (2017), for example).

Moreover, as the random variables $Z_k^{a,b}$, $k \in \mathbb{Z}$, are uniformly bounded by the constant $(|a|+|b|)/p_d$, Eq. (B2), i.e., the classical condition of Ibragimov, implies condition (DMR) in Rio (2017). Then, convergence in Eq. (B1) follows from Theorem 4.2 in Rio (2017).

Applying the Delta method to the differentiable function $(f,g) \mapsto f/g$ yields Eq. (4). □

Acknowledgements The authors thank the data providers of this study, namely the Saxon State Office for the Environment, Agriculture and Geology, the German Weather Service (DWD) and the Movement Data Bank. All data used here are freely available from the data providers. The authors also thank one anonymous referee and the editor Salvatore Grimaldi for their helpful comments that improved this work.

Author Contributions All authors contributed to the study conception and design as well as material preparation, data collection and analysis. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Svenja Fischer has received funding by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in terms of the research unit SPATE (FOR2416, project number FI-2439/3-2). Marco Oesting was partially supported by the project "Climate Change and Extreme Events—ClimXtreme Module B—Statistics (subproject B3.1)" funded by the German Federal Ministry of Education and Research (BMBF) with the grant number 01LP1902I. Alexander Schnurr was supported by the DFG in terms of the project SCHN 1231-3/2.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bandt C (2019) Small order patterns in big time series: a practical guide. Entropy 21(6):613

Bandt C, Pompe B (2002) Permutation entropy: a natural complexity measure for time series. Phys Rev Lett 88(174):102

Betken A, Schnurr A (2023) Depth patterns

Bücher A, Kojadinovic I (2016) Dependent multiplier bootstraps for non-degenerate u-statistics under mixing conditions with applications. J Stat Plan Inference 170:83–105

Bücher A, Ruppert M (2013) Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. J Multivar Anal 116:208–229

Buhrmester V, Münch D, Arens M (2021) Analysis of explainers of black box deep neural networks for computer vision: a survey. Mach Learn Knowl Extr 3(4):966–989

Drees H (2015) Bootstrapping empirical processes of cluster functionals with application to extremograms. arXiv preprint arXiv:1511.00420

Ferreira J, de Sousa Ribeiro M, Gonçalves R, et al (2022) Looking inside the black-box: logic-based explanations for neural networks. In: Proceedings of the international conference on principles of knowledge representation and reasoning, pp 432–442

Finke N, Möller R, Mohr M (2020) Multivariate ordinal patterns for symmetry approximation in dynamic probabilistic relational models. In: AI 2021: advances in artificial intelligence, pp 189–196

Ghada W, Casellas E, Herbinger J et al (2022) Stratiform and convective rain classification using machine learning models and micro rain radar. Remote Sens 14(18):4563

Huang J, Liu F, Xue Y et al (2015) The spatial and temporal analysis of precipitation concentration and dry spell in Ginghai, northwest China. Stoch Environ Res Risk Assess 29:1403–1411

Huard D, Mailhot A, Duchesne S (2010) Bayesian estimation of intensity-duration-frequency curves and of the return period associated to a given rainfall event. Stoch Environ Res Risk Assess 24:337–347

Ibebuchi CC, Abu I (2023) Rainfall variability patterns in Nigeria during the rainy season. Sci Rep 13:7888

Keller K, Lauffer H (2003) Symbolic analysis of high-dimensional time series. Int J Bifurcat Chaos 13(9):2657–2668

Keller K, Sinn M (2010) Kolmogorov-sinai entropy from the ordinal viewpoint. Phys D: Nonlinear Phenom 239(12):997–1000

Liang Y, Li S, Yan C et al (2021) Explaining the black-box model: a survey of local interpretation methods for deep neural networks. Neurocomputing 419:168–182. https://doi.org/10.1016/j.neucom.2020.08.011

Long JA, Nelson TA (2013) A review of quantitative methods for movement data. Int J Geogr Inf Sci 27(2):292–318

Meyer H, Kühnlein M, Appelhans T et al (2016) Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. Atmos Res 169:424–433

Mohr M (2022) Learning from ups and downs: multivariate ordinal pattern representations for time series. Lübeck University, PhD-thesis

Mohr M, Finke N, Möller R (2020) On the behaviour of permutation entropy on fractional Brownian motion in a multivariate setting. In: Proceedings of APSIPA-ASC, pp 189–196

Mohr M, Wilhelm F, Hartwig M, et al (2020) New approaches in ordinal pattern representations for multivariate time series. In: FLAIRS Conference, pp 124–129

Mumuni A, Mumuni F (2022) Data augmentation: a comprehensive survey of modern approaches. Array 16(100):258

Oesting M, Huser R (2022) Patterns in spatio-temporal extremes. arXiv preprint arXiv:2212.11001

Pham TD, Tran LT (1985) Some mixing properties of time series models. Stoch Process Appl 19(2):297–303

Piek AB, Stolz I, Keller K (2019) Algorithmics, possibilities and limits of ordinal pattern based entropies. Entropy 21(6):547

Power R, Venter L, Botha MV, et al (2021) Repatriating leopards into novel landscapes of a South African province. Ecol Solut Eviden e12046

Rio E (2017) Asymptotic theory of weakly dependent random processes. Springer, New York

Schnurr A, Fischer S (2022) Generalized ordinal patterns allowing for ties and their applications in hydrology. Comp Stat Data Anal 171(107):472

Sinn M, Keller K (2011) Estimation of ordinal pattern probabilities in gaussian processes with stationary increments. Comp Stat Data Anal 55:1781–1790

Unakafova V, Keller K (2013) Efficiently measuring complexity on the basis of real-world data. Entropy 15:4392–4415

Wang G (2019) Machine learning for inferring animal behavior from location and movement data. Eco Inform 49:69–76

Weiß CH (2022) Non-parametric tests for serial dependence in time series based on asymptotic implementations of ordinal-pattern statistics. Chaos: Interdiscip J Nonlinear Sci 32(9):093,107

Weiß CH, Schnurr A (2023) Generalized ordinal patterns in discrete-valued time series: non-parametric testing for serial dependence. J Nonparametr Stat

Winterrath T, Brendel C, Hafer M, et al (2018) Radklim version 2016.003: Reprocessed gauge-adjusted radar-data, one-hour precipitation sums (rw)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.